FACE DETECTION IN DIGITAL IMAGES

Technical Field of the Invention

The present invention relates to digital colour images and, in particular, to the detection of faces in colour digital images.

Background Art

Colour digital images are increasingly being stored in multi-media databases, and utilised in various computer applications. In many such applications it is desirable to be able to detect the location of a face in a visual image as one step in a multi-step process. The multi-step process can include content-based image retrieval, personal identification or verification for use with automatic teller machines or security cameras, or automated interaction between humans and computational devices.

Various prior art face detection methods are known including eigenfaces, neural networks, clustering, feature identification and skin colour techniques. Each of these techniques has its strengths and weaknesses, however, one feature which they have in common is that they are computationally intensive and therefore very slow, or they are fast but not sufficiently robust to detect faces.

The eigenface or eigenvector method is particularly suitable for face recognition and there is some tolerance for lighting variation, however it does not cope with different viewpoints of faces and does not handle occlusion of various facial features (such as occurs if a person is wearing sunglasses). Also it is not scale invariant.

The neural network approach utilises training based on a large number of face images and non-face images and has the advantages of being relatively simple to implement, providing some tolerance to the occlusion of facial features and some tolerance to lighting variation. It is also relatively easy to improve the detection rate by re-training the neural network using false detections. However, it is not scale invariant,

does not cope with different viewpoints or orientation, and leads to an exhaustive process to locate faces on an image.

The clustering technique is somewhat similar to the eigenface approach. A pixel window (eg. 20 x 20) is typically moved over the image and the distance between the
5 resulting test pattern and a prototype face image and a prototype non-face image is represented by a vector. The vector captures the similarity and differences between the test pattern and the face model. A neural network can then be trained to classify as to whether the vector represents a face or a non-face. While this method is robust, it does not cope with different scales, different viewpoints or orientations. It leads to an
10 exhaustive approach to locate faces and relies upon assumed parameters.

The feature identification method is based upon searching for potential facial features or groups of facial features such as eyebrows, eyes, nose and mouth. The detection process involves identifying facial features and grouping these features into feature pairs, partial face groups, or face candidates. This process is advantageous in that
15 it is relatively scale invariant, there is no exhaustive searching, it is able to handle the occlusion of some facial features and it is also able to handle different viewpoints and orientations. Its main disadvantages are that there are potentially many false detections and that its performance is very dependent upon the facial feature detection algorithms used.

20 The use of skin colour to detect human faces is described in a paper by Yang J and Waibel A (1995) "Tracking Human Faces in Real-Time" CMU-CS-95-210, School of Computer Science, Carnegie Mellon University. This proposal was based on the concept that the human visual system adapts to different levels of brightness and to different illumination sources which implies that the human perception of colour is consistent
25 within a wide range of environmental lighting conditions. It was therefore thought

possible to remove ●ghtness from the skin colour representation while preserving accurate, but low dimensional, colour information. As a consequence, in this prior art technique, the chromatic colour space was used. Chromatic colours (eg. r and g) can be derived from the RGB values as:

5

$$r = R / (R + G + B) \quad \text{and} \quad g = G / (R + G + B)$$

These chromatic colours are known as "pure" colours in the absence of brightness.

Utilising this colour space, Yang and Waibel found the distribution of skin colour of different people, including both different persons and different races, was

10 clustered together. This means that the skin colours of different people are very close and that the main differences are in differences of intensity.

This prior art method first of all generated a skin colour distribution model using a set of example face images from which skin colour regions were manually selected. Then the test image was converted to the chromatic colour space. Next each image in the

15 test image (as converted) was then compared to the distribution of the skin colour model. Finally, all skin colour pixels so detected were identified, and regions of adjacent skin colour pixels could then be considered potential face candidates.

This prior art method has the advantage that processing colour is much faster than processing individual facial features, that colour is substantially orientation invariant

20 and that it is insensitive to the occlusions of some facial features. The system is also substantially viewpoint invariant and scale invariant. However, the method suffers from a number of disadvantages including that the colour representation of the face can be influenced by different lighting conditions, and that different cameras (eg. digital or film) can produce different colour values even for the same person in the same environment.

However a significant disadvantage of the prior art methods is that the skin colour model is not very discriminating (ie. selecting pixels on a basis of whether they are included in the skin colour distribution results in a lot of non-skin colour pixels being included erroneously). It is also difficult to locate clusters or regions of skin colour pixels

5   that can be considered as candidate faces.

Disclosure of the Invention

An object of the present invention is to provide an improved method of detecting one or more faces in digital colour images.

In accordance with a first aspect of the present invention there is disclosed a

10   method of detecting a face in a colour digital image formed of a plurality of pixels, said method comprising the steps of:

(i)   testing the colour of said pixels to determine those said pixels having predominantly skin colour, said testing utilising at least one image capture condition provided with said image; and

15   (ii)   subjecting only said those pixels determined in step (i) as having predominantly skin colour to further facial feature analysis whereby those said pixels not having a predominantly skin colour are not subjected to said further facial feature analysis.

Preferably each image capture condition is acquired at a time the image is

20   captured. Advantageously, the image is encoded according to a predetermined format and at least one image capture condition is represented as meta-data associated with the format. Most preferably the at least one image capture condition comprises lighting conditions at a time the image was captured.

In a particular implementation step (i), comprises the sub-step, preceding said

25   testing, of:

(a)   dividing said image into a plurality of regions, each said region comprising a plurality of said pixels; and

wherein said testing is performed on pixels within each said region to determine those ones of said regions that are predominantly skin colour and step (ii) comprises performing said further facial feature analysis on only those said regions determined to be predominantly of skin colour.

In accordance with another aspect of the present invention, there is disclosed a method of detecting a face in a colour digital image, said method comprising the steps of:

(i)   segmenting said image into a plurality of regions each having a substantially homogenous colour;

(ii)   testing the colour of each said region created in step (i) to determine those regions having predominantly skin colour; and

(iii)   subjecting only the regions determined in step (ii) to further facial feature analysis whereby said regions created in step (i) not having a predominantly skin colour are not subjected to said further feature analysis.

Apparatus and computer readable media for performing the invention are also disclosed

Brief Description of the Drawings

A number of embodiments of the present invention will now be described with reference to the drawings in which:

Fig. 1 is a schematic representation of the pixels of a colour digital image;

Fig. 2 shows the segmenting of the image of Fig. 1 into a plurality of regions each having a substantially homogenous colour according to a first embodiment;

Fig. 3 is a flow chart of a face detection process according to the first embodiment;

Fig. 4 is a schematic block diagram of a general purpose computer upon which embodiments of the present invention can be practised;

Fig. 5 is a flow chart depicting the generation of a face colour distribution model;

Fig. 6 is a flow chart of a face detection process according to a second embodiment; and

Fig. 7 is a flow chart of a face detection process according to a third embodiment.

Detailed Description including Best Mode

Fig. 1 illustrates a typical colour digital image 1 having a size of 832 x 624 pixels 5, each of which has an RGB value.

According to a first embodiment of the present invention, rather than consider the skin colour of the image on a pixel by pixel basis as described above in relation to the prior art of Yang and Waibel, the image 1 is segmented into a number of regions. An example of such segmentation is schematically illustrated in Fig. 2 on the segmentation basis that all the pixels in each region 2 have substantially the same colour. Alternatively the image may be segmented into arbitrary regions for processing.

The first embodiment implements a process 30 illustrated in the flow chart of Fig. 3, in which the regional segmentation of the image is carried out at step 31. Next, the regions of the image are converted in step 32 into the chromatic colour space (as described above). The next step 33 is to select those of the regions determined in step 31 which have a specified percentage (typically in the range 90-95%) of pixels having a skin colour. These selected regions are then conveniently represented by a boundary box or other boundary indication. Finally, at step 34 the selected regions, including any combinations of overlapping regions, are subjected to a further analysis (preferably not based on skin colour) to determine if the selected region(s) represent a face or faces.

This initial colour grouping can use any region based colour image segmentation technique. Preferably the image is partitioned into colour regions by seeking connected groups of pixels which have similar colours over a local region. Very small isolated initial spatial groupings can be ignored in order to find major colour regions and reduce a noise effect. A representative colour of each initial spatial region is determined by an average colour value of the region.

A colour region starts from an arbitrarily chosen pixel, which is compared with its neighbouring pixels. The region size is increased by adding neighbouring pixels, which are similar in colour, using a colour similarity threshold T. A neighbouring pixel is added to the region if $|Rp-Rm|<T$ and $|Gp-Gm|<T$ and $|Bp-Bm|<T$, where Rp, Gp, Bp are R, G, B values of the neighbouring pixel and Rm, Gm, Bm represented average R, G, B values of the region.

When a region has no more neighbouring pixels of similar colour, the region stops growing and represents one of the initial spatial groupings. If the region size is below a predetermined threshold value, it is ignored. A region having a pixel number equal or greater than the predetermined threshold is represented by its average colour.

A new pixel which does not yet belong to any region is chosen to start a new colour region. The process continues until each pixel in the image either belongs to an initial spatial grouping or has been ignored as being part of a small region.

The initial spatial groupings provide a colour region segmentation of the image with each region being represented by its average colour.

In this way, for most images where the bulk of the image is not a face, or part of a face, the majority of pixels will be grouped into regions or objects (be they foreground or background, etc) which are clearly not faces. Therefore these non-facial objects can be quickly eliminated on the basis of their colour.

Once the regions have been determined, they are then converted into the "pure" chromatic colour space utilising the equations given above so as to provide r and g values. A generous rule such as a rule that at least 85% of the pixels within a given region be of face colour can be used to select those regions worthy of further examination. Preferably,

5 the test for face colour takes into account the nature of the original image, for example whether the image was taken with or without a flash. This information can be determined from the image source, eg. a camera.

Thereafter, only those selected regions are subjected to a further test to determine the presence of facial features. This further test provides a conclusive determination as to

10 whether or not a region constitutes a face. In this connection, the further test is likely to be computationally slower and therefore the above described elimination of regions ensures that the computationally slow method is only applied to relatively small portions of the overall image. Thus the total processing time is reduced. Accordingly, the above method performs a computationally simple process on most, if not all pixels, and then

15 only performs complex examination on skin colour regions.

The preferred method of verifying if a region represents a face relies upon edge detection techniques as a means of detecting facial features. In particular facial features such as eyes, eyebrows and mouths often appear as dark bars on a face and thus provide dark edges.

20 The preferred form of edge detection is use of an edge detection filter. This utilises two functions operating in orthogonal directions. To detect a horizontal bar a second derivative Gaussian function is used in the vertical direction and a Gaussian function is used in the horizontal direction.

Once an edge has been determined in this way each detected edge is examined.

25 Any pair of detected edges can be found to be derived from, and thus be indicative of, a

pair of eyes, a pair ⬤ eyebrows, or an eye and associated eyebrow, depending upon the relative position and size of the detected edges. Similarly, an individual edge can be derived from, and thus be indicative of, a mouth if it is located at an appropriate position relative to the eyes and/or eyebrows already detected.

By proceeding in this fashion, a given region begins to accumulate facial features building from skin tone through eyebrows/eyes and then to a mouth. The more facial features found for a given region which is a face candidate, the greater the possibility that the candidate actually is a face.

Furthermore, the above described method has the advantage that it is able to cater for the circumstance where a face is backgrounded against a background region of substantially the same colour. Under these circumstances, in Yang and Waibel's method, no boundary between the face and the background would be likely to be detected. Therefore the region as a whole would be selected for further testing. However, the above method utilises the full colour space to segment the image, before making decisions about which pixels are skin colour. Consequently the face is more likely to be separated from the background. In addition, the method is naturally independent of orientation or partial occlusion of the face.

Furthermore, the above method also permits false positives to be examined at the further stage and therefore does not exclude from subsequent testing regions which are likely to be ultimately determined as a facial region.

The first embodiment described above notes that the nature of the original image may be taken into account when performing an initial face detection process. Further embodiments to be now described build upon this feature.

When an image is captured using a camera it is necessary either for the person taking the picture to manually establish the camera settings (such as shutter speed,

aperture, focal length etc), or for the camera to perform this operation automatically. Whichever is the case, the settings of the camera directly effect the appearance and quality of the image taken. In particular, the perceived brightness, colour, and sharpness of the objects within an image all depend on how the settings of the camera are

5   configured. For example, it is possible to take two pictures of the same scene with different camera settings and to obtain two images in which the same objects appear with different colours and brightness. Therefore, the ability to calibrate (in particular) colour information contained in (digital) images enables a broad variety of object detection and classification tasks in which colour is a strong discriminating feature.

10   Face detection is one such example application, and the present inventors have determined that the creation of face colour distribution models (CDM's), each adapted to specific lighting conditions, that can improve both the accuracy and reliability of face detection. Variations in lighting conditions can result from the use of a flash, such being a feature recognised as contributing in the face detection method of the first embodiment.

15   Since lightness is representative of colour features such as luminance and chrominance, such features may be used to quantify face detection.

Before an image can be processed using a face colour distribution model, the face colour distribution model must be constructed. This is performed according to a method 50 shown in Fig. 5. The method 50 firstly gathers image samples at step 52 that

20   are representative images that contain faces, the images being acquired under a variety of lighting conditions and thus indicative of changes in luminance and chrominance. These images are then manually examined in step 54 to extract regions of skin colour for further processing in model formation. Step 54 may be performed by manually drawing a bounding box around a sample of face coloured pixels. Step 56, which follows, derives

25   colour representation values for the extracted pixels. This may be performed by

transforming the e⬤ted pixels into a perceptual colour s⬤ce such as CIE L*u*v or CIE L*a*b, so that each pixel is represented by at least a 2-dimensional vector. Alternatively other colour spaces such as HLS and HSV may be used. Preferably each pixel is represented as a length-3 vector incorporating both the luminance and
5   chrominance values.

The colour representation values of pixels are then divided at step 58 into a number of sets (58a, 58b ... 58n) according to the lighting conditions present when each of the images were captured. Example sets are flash, non-flash, indoor, outdoor, and combinations of these. Alternatively, lighting parameters obtained directly from the
10   camera such as the operation of a flash, may be used to identify and distinguish the sets. Other lighting conditions such as bright or cloudy, dusk or dawn, or a type of artificial light such as fluorescent, incandescent or halogen, may be used or detected for these purposes. These details may be provided by means of human input at the time of image capture.

15   For each of the sets (58a ... 58n) of face samples, step 60 then constructs a corresponding colour distribution model (CDM) (60a ... 60n) that best fits the samples of face colour pixels. The CDM can be a histogram, a probability density function, or a binary map. In one embodiment, a mixture of Gaussian PDF's are fit to the sample data using techniques known in the art such as the expectation maximisation (EM) algorithm,
20   with either cross-validation, jackknife, and bootstrap techniques being used to estimate the goodness of fit of the model.

When each CDM (60a ... 60n) has been constructed, it is then desirable as shown in step 62 to establish a corresponding probability threshold (62a ... 62n) below which a colour vector is to be classified as relating to a non-face pixel, and above which
25   the colour vector is to be classified as a potential face pixel. Additionally, the face colour

probability can b●●ed directly in further facial image analysis steps detailed below. In the preferred embodiment, the CDM is constructed from colour representation values derived using a perceptual colour space (such as CIE L*u*v or CIE L*a*b) and then transformed back into the colour format of the input image, ie., either RGB or YUV. This

5   removes the necessity for transforming the input image into the perceptual colour space.

Since different image capture devices have differing performance, often determined by the quality and size of optical components (eg. lens, mirrors, aperture etc.), typically a CDM or a set of CDM's are generated for a particular capture device. In one implementation, where the image capture device (eg. camera) includes a light meter, a

10   reading from the light meter at the moment the image was captured can be used to determine the required CDM. In this fashion, a greater range of colour models may be devised and can be selected without possible human interference. Such interference may occur where the human user manually selects the operation of the flash where otherwise automatic operation of the flash would not be required. Further, the flash/outdoors

15   example sets above give rise to four (4) sets of CDM's. Using a light meter with, say, 4-bit encoding, can provide sixteen (16) models. Also, use of a light meter provides for enhanced reproducability of results and enables the face samples used to generate the models to be taken under laboratory conditions and installation at the time of camera manufacture.

20   The processing 70 of an image according to a second embodiment is shown in Fig. 6. An input image is provided at step 72 and at step 74 the lighting conditions under which the image was captured are determined. Such a determination may be based on binary data obtained directly from the camera (eg. flash+indoors, no_flash+outdoors, no_flash+indoors, flash+outdoors) or corresponding meta-data provided with, or

25   accompanying, the image, which may be encoded or otherwise communicated according

to a predetermined format. Once the lighting conditions are determined, a corresponding or closest CDM is selected from a bank of look-up tables 78 retaining the CDM's (60a ... 60n) previously determined. At step 80, a first pixel of the input image 72 is selected and at step 82 is tested to see if the (RGB or YUV) colour of the pixel is contained in the

5   selected CDM (60a ... 60n).

The steps shown in Fig. 6 following the comparison of step 82 depend upon the manner in which the CDM's are stored. In a preferred implementation, the selected thresholds of step 62 (Fig. 5) are used to construct a binary map or look-up table, where a representative colour vector is represented by a 1 if it is a colour vector that is contained

10   within the thresholded face colour distribution, and by a 0 if the colour vector does not occur in the thresholded colour distribution. Alternatively, the CDM may represent the frequencies of the representative colour vectors of the thresholded colour distribution (ie. the CDM is effectively a histogram of representative colour vectors). A further variation is the case where a sampled distribution is approximated by a parametric model such as a

15   Gaussian or a mixture of Gaussians. In the latter case, the CDM comprises the parameters of the model (eg. mean, co-variance)

As seen in Fig. 6, and according to the preferred implementation, a 1 or 0 value arising from step 82 is added to a map in step 84. Step 86 determines if there are more pixels in the image to be processed and step 88 obtains and passes the next pixel to

20   step 82 for appropriate testing. When all pixels have been tested against the selected CDM, step 90 indicates the result of the preceding steps as being a binary face image map formed using detected skin-coloured pixels.

The map is then subjected at step 92 to further analysis of the skin-coloured pixels to provide at step 94 a face detection map for the image. The further analysis of

step 92 is, like the first embodiment, preferably independent of considerations of facial colour.

In practice, binary face map formed at step 90 may contain areas where either there are small non-face pixels (0's) surrounded by face pixels (1's), or vice versa. One approach for further analysis according to step 92 is processing the binary face image map so as to set to 0 any pixel locations which are contained in areas that are smaller than the smallest size of a potential face and to set any 0 pixel locations to 1 if they are surrounded by likely face colour pixels. This may be performed using a pair of morphological opening and closing operations with suitably shaped structuring elements. A first structuring element such as:

$$\begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix},$$

is used in the opening operation to remove potential face candidate pixel locations below this size. A second structuring element such as:

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix},$$

is used in the closing operation to fill any holes in potential face candidate pixel locations.

Alternative approaches to the use of the structuring elements include using a Hough transform, or to count the number of pixels in the region having skin colour and to threshold that count against a predetermined percentage value. Other methods may be used to perform these tasks.

The result of the process 70 of Fig. 6 is a face detection map of pixel locations in the input image at which a face has been detected, and in all likelihood, a face is present.

The aforementioned edge detection method of further processing the likely face pixels to determine if a face exits may then be performed on the face detection map 94 resulting from the method 70.

In the preferred embodiment, the face colour distribution models are built for a number of separate lighting conditions, such as flash, non-flash, indoor, outdoor, and the like. However, this technique may be extended to the more general case of arbitrary lighting conditions based directly on parameters obtained from the camera. A list of camera parameters that may be of use in this situation is as follows:

(i)   white balance;

(ii)   white balance mode

(iii)   aperture (iris);

(iv)   shutter speed;

(v)   auto gain control (AGC);

(vi)   auto exposure (AE) mode;

(vii)   gamma;

(viii) pedestal level; and

(ix)   flare compensation.

The parameters obtained from the camera are preferably obtained from a meta-data stream associated with the capture of each image (or video sequence). Examples, of such transmission protocols include IEEE 1394 ("firewire"). Also the ISO standards have defined methods for attaching meta-data to images and video in MPEG-7, MPEG4, and JPEG.

Whilst the first embodiment described with reference to Figs. 1 to 3 divides the image according to regions of substantially homogeneous colour, the second embodiment, and a third embodiment, are not so constrained.

The third embodiment is depicted in Fig. 7 by a method 150 where an input image 152 is provided and processed according to steps 154, 156 and 158 corresponding to steps 74, 76 and 78, respectively, of the second embodiment. Once the appropriate CDM has been selected in step 156, step 160 follows to process the input image as one or

5   more regions. As a single region, the entirety of the image is processed on a pixel-by-pixel basis. Alternatively, the input image may be geometrically divided into simple pixel blocks (eg. 25 x 25 pixels, 10 x 20 pixels) which can be formed and processed in raster order. A further alternative, is where the regions, like the first embodiment, are divided on the basis of substantially homogeneous colour.

10   Step 162 selects a first region to be processed and step 164 a first pixel of that region. Step 166 compares the selected pixel with the CDM in a manner corresponding to step 82 of the second embodiment. Where the pixel matches the model, step 168 increments a count of pixels in the region meeting that criteria. Step 170 determines whether there are any other pixels is the region to be processed and , if so, step 172

15   obtains the next pixel and returns to step 166 for appropriate testing. When all pixels in the region have been processed, step 174 follows to compare the percentage of pixels classified for the region as skin colour against a predetermined percentage threshold value. Where the percentage is less than the predetermined number, the region is considered a non-face region and step 176 follows to test if there are any more regions to

20   be processed. If so, step 178 selects the next region and returns processing to step 164. The count is then re-set. If not, the method 150 ends at step 184.

Where the percentage exceeds the predetermined percentage, the region is considered a possible face region and step 180 follows to assess the region according to further facial detection analysis. Where such analysis does not detect a face, the

25   method 150 proceeds to step 176 to process any further regions. Where the further

analysis of step 180 detects a face, step 182 registers that region as a face region and returns to step 176.

An example of the further analysis able to be performed as a consequence of appropriate marking at step 180 is the edge detection method described above in relation
5 to the first embodiment.

The above described embodiments each indicate that face detection in images may be performed as a two stage process, a first representing something akin to a first filtering of the image to obtain likely candidate pixels or regions, and the second representing more thorough analysis to provide an actual determination on those pixels or
10 regions passed by the first stage. In each case, lighting conditions associated with the capture of the image contribute to the determination performed by the first stage.

The above described methods are preferably practiced using a conventional general-purpose computer system 100, such as that shown in Fig. 4 where the processes of Fig. 3 and/or Figs. 5 and 6 are implemented as software, such as an application
15 program executing within the computer system 100. In particular, the steps of the methods are effected by instructions in the software that are carried out by the computer. The software may be divided into two separate parts; one part for carrying out the above method steps; and another part to manage the user interface between the latter and the user. The software may be stored in a computer readable medium, including the storage
20 devices described below, for example. The software is loaded into the computer from the computer readable medium, and then executed by the computer. A computer readable medium having such software or computer program recorded on it is a computer program product. The use of the computer program product in the computer preferably effects an advantageous apparatus for detecting face candidate regions in accordance with the
25 embodiments of the invention.

The computer system 100 comprises a computer module 101, input devices such as a keyboard 102 and mouse 103, output devices including a printer 115 and a display device 114. A Modulator-Demodulator (Modem) transceiver device 116 is used by the computer module 101 for communicating to and from a communications network 120, for

5   example connectable via a telephone line 121 or other functional medium. The modem 116 can be used to obtain access to the Internet, and other network systems, such as a Local Area Network (LAN) or a Wide Area Network (WAN), these being possible sources of input images and destinations for detected faces.

The computer module 101 typically includes at least one processor unit 105, a

10   memory unit 106, for example formed from semiconductor random access memory (RAM) and read only memory (ROM), input/output (I/O) interfaces including a video interface 107, and an I/O interface 113 for the keyboard 102 and mouse 103 and optionally a joystick (not illustrated), and an interface 108 for the modem 116. A storage device 109 is provided and typically includes a hard disk drive 110 and a floppy disk

15   drive 111. A magnetic tape drive (not illustrated) may also be used. A CD-ROM drive 112 is typically provided as a non-volatile source of data. The components 105 to 113 of the computer module 101, typically communicate via an interconnected bus 104 and in a manner which results in a conventional mode of operation of the computer system 100 known to those in the relevant art. Examples of computers on which the

20   embodiments can be practised include IBM-PC's and compatibles, Sun Sparcstations or alike computer systems evolved therefrom.

Typically, the application program of the preferred embodiment is resident on the hard disk drive 110 and read and controlled in its execution by the processor 105. Intermediate storage of the program and any data fetched from the network 120 may be

25   accomplished using the semiconductor memory 106, possibly in concert with the hard

disk drive 110. In some instances, the application program may be supplied to the user encoded on a CD-ROM or floppy disk and read via the corresponding drive 112 or 111, or alternatively may be read by the user from the network 120 via the modem device 116. Still further, the software can also be loaded into the computer system 100 from other

5   computer readable medium including magnetic tape, a ROM or integrated circuit, a magneto-optical disk, a radio or infra-red transmission channel between the computer module 101 and another device, a computer readable card such as a PCMCIA card, and the Internet and Intranets including e-mail transmissions and information recorded on Websites and the like. The foregoing is merely exemplary of relevant computer readable

10   mediums. Other computer readable mediums may be practiced without departing from the scope and spirit of the invention.

The further processing to candidate face images and regions may also be performed by or using the computer system 100 and known arrangements for such processing.

15   The method of detecting face candidate regions may alternatively be implemented in dedicated hardware such as one or more integrated circuits performing the functions or sub functions of Fig. 3 and/or Figs. 5 and 6. Such dedicated hardware may include graphic processors, digital signal processors, or one or more microprocessors and associated memories.

20   ## Industrial Applicability

It is apparent from the above that the embodiments of the invention are applicable in fields such as content-based image retrieval, personal identification or verification for use with automatic teller machines or security cameras, or automated interaction between humans and computational devices.

The foregoing describes only some embodiments of the present invention, and modifications and/or changes can be made thereto without departing from the scope and spirit of the invention as defined in the claims.